

CLASSIFICADORES SEMÂNTICOS NA IDENTIFICAÇÃO DE EQUIVALENTES INTERLINGUAIS PARA EXPRESSÕES IDIOMÁTICAS: DESCASCANDO ESSE ABACAXI

Semantic Classifiers to Identify Equivalent Idioms across Languages: a Tough Nut to Crack

DOI: 10.14393/LL63-v41S-20226-01

Julia de Souza BATISTA*

Isabela MOREIRA DE OLIVEIRA**

Stella Esther Ortweiler TAGNIN***

Elisa Duarte TEIXEIRA****

Rozane Rodrigues REBECHI*****

* Bacharel em Línguas Estrangeiras Aplicadas (UnB). Graduanda em Tradução – Inglês pela Universidade de Brasília. ORCID: 0009-0007-7282-4999. E-mail para contato: juliabatista.lea(AT)gmail.com.

** Mestra em Estudos da Tradução (UnB). Doutoranda do PPGET, Universidade Federal do Rio Grande do Sul. ORCID: 0009-0003-3796-8787. amoreirabela(AT)gmail.com.

*** Livre-Docente em Estudos Linguísticos e Literários em Inglês (USP), Professora Sênior atuando em dois programas de pós-graduação na USP: Estudos Linguísticos e Literários em Inglês (PPGELLI) e Línguas Estrangeiras e Tradução (LETRA). ORCID: 0000-0002-5517-2710. seotagni(AT)usp.br.

**** Doutora em Estudos Linguísticos e Literários em Inglês (USP). Professora de Tradução – Inglês do Departamento de Línguas Estrangeiras e Tradução da Universidade de Brasília. ORCID: 0000-0003-3472-3605. E-mail para contato: elisadut(AT)unb.br.

***** Doutora em Estudos Linguísticos e Literários em Inglês (USP). Universidade Federal do Rio Grande do Sul. ORCID: 0000-0002-1878-7548. E-mail par contato: rozane.rebechi(AT)ufrgs.br.

RESUMO: Expressões idiomáticas (EIs), como “a cereja do bolo”, muito utilizadas em interações orais e escritas do dia a dia, apresentam desafios para a tradução e o aprendizado de línguas devido à não transparência de seu significado e por serem geralmente assimiladas por exposição direta. Em vista disso, este estudo visa contribuir para o objetivo geral do projeto de pesquisa ao qual se filia: desenvolver um dicionário onomasiológico multilíngue de EIs contendo vocabulário relacionado à alimentação. Para tanto, apresenta uma metodologia de elaboração e testagem inicial de um sistema de descritores semânticos, a serem utilizados, em etapa futura, na identificação semiautomática de equivalentes interlinguais e relações de conteúdo intralinguais entre EIs do banco de dados do projeto. Testes de atribuição de descritores semânticos a EIs por anotadores humanos, em português e inglês, revelaram baixo índice de consenso, o que indica desafios práticos e conceituais a serem solucionados na continuidade da pesquisa.

PALAVRAS-CHAVE: Expressões idiomáticas. Lexicografia. Fraseologia. Classificação semântica. PLN.

ABSTRACT: Idiomatic expressions (IEs), such “the cherry on top,” commonly used in everyday oral and written interactions, pose challenges for translation and language learning due to the opacity of their meanings and the fact that their acquisition is primarily through direct exposure. On that account, this study aims to contribute to a larger research project focusing on the development of a multilingual, onomasiological dictionary of IEs featuring food-related vocabulary by presenting a methodology for the initial design and testing of a system of semantic classifiers. The main project envisions the future use of these classifiers in the semi-automatic identification of interlingual equivalents and intralingual content relations among IEs of the project’s database. Tests involving the assignment of semantic classifiers to IEs in Portuguese and English by human annotators revealed a low level of agreement, indicating practical and conceptual challenges to be addressed throughout the continuation of the project.

KEYWORDS: Idiomatic Expressions. Lexicography. Phraseology. Semantic classification. NLP.

1 Introdução

Expressões idiomáticas (EIs), como “chorar as pitangas”, “a cereja do bolo” e “bananeira que já deu cacho”, são construções linguísticas com raízes culturais profundas e de uso frequente no dia a dia dos falantes de qualquer língua, especialmente em sua modalidade oral (Cowie; Mackin, 1975; Gross, 1988). Por isso, são tema de grande interesse não apenas para professores, estudantes e pesquisadores das áreas de Letras, mas para quaisquer pessoas interessadas em adquirir mais proficiência e conhecimento cultural em seu próprio idioma ou em um idioma estrangeiro (Moreira de Oliveira; Teixeira, 2025).

Conforme explica Tagnin (2005), a **idiomaticidade** ocorre quando o significado de uma expressão é opaco, não transparente – ou seja, seu sentido não pode ser depreendido da somatória dos significados de seus componentes (como “chorar” e “pitangas”, no primeiro exemplo do parágrafo anterior, que, isoladamente, não oferecem qualquer vislumbre sobre o real sentido da expressão: “reclamar, lamentar”). A autora explica que esse fenômeno é chamado de ‘não composicionalidade’.

Tagnin (2005) também aponta que as EIs são caracterizadas por um outro aspecto, a **convencionalidade**, ou seja, aquilo que é consolidado pelo uso numa determinada língua ou comunidade linguística, o que explicaria, entre outras coisas, a diferença entre os sentidos atribuídos a combinações de palavras de uma mesma língua em países diferentes, ou por gerações ou regiões de um mesmo país. Em Portugal, por exemplo, a expressão “ter a burra nas couves” significa “ter um grande problema”, mas a combinação não possui significado algum no Brasil.

Essas duas características das EIs fazem com que seu reconhecimento e compreensão dependam de uma exposição prévia (Leffa, 2000; Mattos, 2003), seja por meio de sua ocorrência natural no discurso, seja pelo contato direto com repertórios previamente elaborados, como listas, materiais didáticos ou dicionários. É a partir dessa exposição que os usuários e aprendizes da língua serão capazes de desenvolver uma competência lexical e cultural para compreender seu emprego e, posteriormente, utilizá-las adequadamente em sua produção escrita e/ou oral (Tagnin, 2005; Xatara; Succi, 2008; Termignoni; Finatto, 2017).

EIs também são um entrave para o Processamento da Linguagem Natural (PLN) (Sag *et al.*, 2002), especialmente do ponto de vista multilíngue, impactando o desempenho de

aplicações linguísticas diversas, como tradutores automáticos, sistemas de produção e reconhecimento de fala e texto, de desambiguação de sentido e de recuperação de informação (Adewumi *et al.*, 2022; Korkontzelos *et al.*, 2013).

Nas últimas décadas, desafios inerentes à perspectiva multilíngue e contrastiva das Els vêm sendo amplamente descritos por pesquisadores da Tradução, como Newmark (1988), Tagnin (1988), Baker (1992), Xatara (1998), Xatara; Riva; Rios (2001), Tagnin (2005), Simão (2011), Silva; Teixeira (2021) e Rebechi; Trindade (2021), entre tantos outros. Essa perspectiva multilíngue é o foco deste trabalho, que visa compilar uma lista de classificadores semânticos e testá-los na identificação de equivalências interlinguais e relações intralinguais entre as Els contendo palavras relacionadas à alimentação. O trabalho é parte do projeto de pesquisa *Expressões Idiomáticas com a Temática Alimentação: uma proposta onomasiológica*, doravante “Projeto”, em desenvolvimento pelo grupo de pesquisa TerminiTraDiCo – Terminologia e Tradução Direcionadas por Corpus – e colaboradores. Nele, propõe-se a criação de uma plataforma inovadora, interativa e de acesso livre para consulta a um repertório multilíngue de Els estruturado onomasiologicamente, isto é, um sistema em que o usuário possa buscar uma EI também pelos seus sentidos. Por exemplo, se buscar “problema”, encontrará “ALGO ser um abacaxi” ou “ALGO ser um pepino”; se quiser expressar “facilidade”, encontrará “ALGO ser mamão com açúcar” ou “ALGO estar no papo”, podendo fazer o mesmo em todas as línguas contempladas no projeto – por enquanto, português brasileiro (PB), inglês (EN), chinês (ZH), espanhol (ES) e italiano (IT).

O Projeto é uma ampliação do que foi proposto inicialmente por Silva e Teixeira (2021), Moreira de Oliveira (2022) e Moreira de Oliveira e Teixeira (2025), mas envolve a coleta de novas Els em PB e EN e a inclusão dos outros três idiomas mencionados acima, por agora. Para tanto, utiliza-se um modelo de ficha lexicográfica e um banco de dados atualizados. O objetivo da obra é servir de fonte de consulta para tradutores, aprendizes de idiomas e até mesmo falantes nativos, além de oferecer informações ontológicas para uso futuro em PLN e Inteligência Artificial (IA).

Este artigo trata das cinco etapas iniciais propostas para viabilizar a construção desse material de referência ampliado, a saber: i) organização das novas Els em PB e EN coletadas pela equipe em categorias relacionadas aos alimentos e a seu preparo (como “frutas”,

“proteínas animais”, “utensílios”, “bebidas”, “grãos”, “vegetais” etc.); ii) elaboração de uma lista padronizada de descritores semânticos (como “facilidade”, “excelência”, “problema”, “exagero” etc.), baseada em materiais lexicográficos existentes e trabalhos anteriores publicados sobre o tema; iii) classificação, por anotadores humanos membros do projeto, das Els em PB e EN anotadas com a categoria “frutas” utilizando esses classificadores; iv) avaliação dos resultados obtidos com relação à eficácia do método de classificação e ao uso dos classificadores na identificação de possíveis sinonímias e equivalências entre as Els em PB e EN; e v) planejamento dos próximos passos da pesquisa. O objetivo principal deste balão de ensaio é refletir sobre a aplicabilidade da metodologia e identificar possíveis entraves e ajustes necessários para que se possa chegar a uma lista completa de classificadores e a um instrumento de categorização e validação aplicáveis à totalidade das Els do banco de dados. A hipótese é de que um sistema adequado de descritores ajudará a identificar equivalentes interlinguais (como “pôr o pão na mesa” e “*bring home the bacon*”) e relações de conteúdo intralinguais (p. ex., “resolver um pepino” e “descascar um abacaxi”).

O artigo está organizado da seguinte maneira: primeiramente, apresenta-se a fundamentação teórica, a partir de um brevíssimo histórico do estudo das expressões idiomáticas na literatura da área, com foco na tradução, nas línguas (PB e EN) e na temática escolhidas para este trabalho inicial. Em seguida, são abordadas as peculiaridades das obras lexicográficas onomasiológicas, das perspectivas teórica e prática, citando exemplos de trabalhos anteriores que utilizaram abordagens semelhantes. Na sequência, apresentamos os fundamentos e a implementação da anotação linguística das Els, necessária para a criação de uma obra onomasiológica. Por fim, descrevemos a metodologia empregada neste trabalho e apresentamos os resultados até agora alcançados, discutindo seus impactos na continuação do Projeto.

2 Expressões idiomáticas: abordagens teóricas e práticas

2.1 Brevíssimo histórico e definições

Uma expressão idiomática, de acordo com Gramley, Gramley e Pátzold (2003, p. 55), é uma “unidade lexical complexa maior do que uma palavra, mas menor do que uma sentença, cujo significado não pode ser apreendido pelo conhecimento do significado de seus

elementos.”¹. Essa definição, adotada pelo Projeto, coincide com os conceitos de **convencionalidade** e **idiomaticidade** apontados por Tagnin (2005) como aspectos indispensáveis à caracterização das Els: convencionalidade sendo a consolidação pelo uso em dada comunidade linguística, e idiomaticidade, a opacidade ou não composicionalidade do sentido a partir do significado de seus itens lexicais.

As Els, assim como outros tipos de expressões fixas, fazem parte do discurso desde tempos imemoriais e são muito mais usadas na produção textual do que as combinações livres (Gross, 1988). Os primeiros registros remontam à epopeia de Gilgamesh, de 2.000 a.C. (Ramos-Nogueira, 2017). Mas o marco inicial dos estudos fraseológicos é atribuído ao lançamento, em 1905 e 1909, de duas obras de Charles Bally, considerado por muitos o fundador da Fraseologia, por suas pesquisas acerca da fixação e dos resultados do uso recorrente de determinadas construções (Ortíz Alvarez, 2000; Welker, 2004). Conforme aponta Corpas Pastor (1996), a Fraseologia formalizou-se como disciplina científica na antiga URSS, na década de 1940. Dentre os estudiosos, a autora destaca o nome de Vinogradov, por ter esclarecido muitas questões não discutidas até então sobre o tema, como a classificação semântica das combinações, e por ter cunhado o termo “unidade fraseológica” (UF), usado na área até hoje.

Outros estudos fundamentais para as primeiras reflexões sobre a Fraseologia foram realizados por pesquisadores como Saussure, Bréal, Greimas, Coseriu, Fillmore, Tristá Pérez, Zuluaga-Ospina e Casares Sánchez, para citar apenas alguns (vide Ramos-Nogueira, 2017, p. 26-30), além de importantes contribuições mais contemporâneas, como as de Mel’cûk (1988), Cowie (1998), Gries (2008) e Polguère (2014). Cada qual abordou alguma questão importante para compreender os problemas característicos da Fraseologia e da Paremiologia, área correlata, como a relação entre as locuções e as palavras compostas, a tipologia das expressões, a concepção de idiomaticidade, a fixidez das unidades, a transformação do plano literal para o idiomático, entre outras questões.

No Brasil, o primeiro compilado de fraseologismos de que se tem notícia é de 1848, de autoria de Perestrelo da Câmara, com o título *Collecção de proverbios, adagios, rifãos anexins sentenças moraes e idiotismos da lingoa portugueza*. Em 1972, Oswaldo Serpa publicou um

¹No original: “a complex lexical item which is longer than a word form but shorter than a sentence and which has a meaning that cannot be derived from the knowledge of its component parts.”

Dicionário de expressões idiomáticas, com foco no par de línguas português-inglês. Em 1989, de uma perspectiva mais acadêmica e teórica, Tagnin abordou o tema em *Expressões idiomáticas e convencionais*, também contrastando português e inglês. Em 2005, a autora lançou a primeira edição de *O Jeito que a gente diz*, relançado em 2013, com alterações substanciais, no sentido de trazer as contribuições inestimáveis da Linguística de Corpus para a área, além de diversos exemplos de Els em outras línguas modernas.

Com o olhar voltado para a língua francesa, Claudia Xatara defendeu seu mestrado sobre o tema em 1994 e, a partir daí, orientou várias pesquisas sobre Fraseologia de uma perspectiva mais onomasiológica. Huélinton Cassiano Riva, um de seus orientandos, defendeu seu doutorado em 2009 (Riva, 2009). Segundo o resumo do trabalho, foi o primeiro a propor um modelo de dicionário de expressões idiomáticas do português do Brasil sob uma perspectiva onomasiológica. Ortíz Alvarez, pesquisadora cubana pioneira da área, radicada no Brasil, defendeu sua tese sobre Els cubanas e brasileiras em 2000 (Ortiz Alvarez, 2000). Contribuiu com a Fraseologia e a Paremiologia, publicando diversos artigos e livros e orientando dezenas de trabalhos na área. Ramos-Nogueira, um de seus orientandos, cunhou o termo “(re)enunciação fraseológica” ao tratar da tradução de fraseologismos em obras literárias no par linguístico espanhol-português (Ramos-Nogueira, 2017). No par de línguas português-italiano, destacam-se nomes como o de Cláudia Zavaglia, da UNESP. De uma perspectiva mais multilíngue, o único dicionário trilingue encontrado na literatura incluindo o português brasileiro foi o *Dicionário de Provérbios Francês - Português - Inglês* (Lacerda; Lacerda; Abreu, 1999).

No que diz respeito ao estudo das Els da área da alimentação, o tema tem sido objeto de algumas pesquisas acadêmicas, tais como Plantin (2011), Seco (2017), Pinnavaia (2018), Silva e Teixeira (2021), Rebechi e Trindade (2021), Moreira de Oliveira (2022), Seco (2024), Seco e Parreira (2025) e Moreira de Oliveira e Teixeira (2025), e de inventários lexicográficos, como os de Jacobson (1993), Palmatier (2000) e Bhala (2009). No entanto, salvo engano, ainda não foi criado nenhum material de referência multilíngue, em papel ou informatizado, com foco nas expressões idiomáticas desse campo semântico específico, e muito menos de uma perspectiva onomasiológica, que seria o formato de organização ideal para contemplar a produção textual em língua materna ou estrangeira, conforme se argumenta na seção a seguir.

2.2 Dificuldades na tradução e no aprendizado das expressões idiomáticas

Els são um exemplo vívido da forte conexão entre língua, cultura e história. No Brasil, por exemplo, a expressão “ALGUÉM SER o rei da cocada preta” é utilizada para se referir de forma crítica e jocosa a uma pessoa arrogante, que se acha muito importante ou superior. A origem dessa El, de acordo com Deonísio da Silva (*apud* Oliveira, 2017), está relacionada à vinda da Família Real portuguesa para o Brasil. Segundo relatos, o rei Dom João VI tinha o privilégio de comer as primeiras cocadas pretas, as mais cobiçadas e, somente depois, os outros nobres podiam se servir (Oliveira, 2017). Esse dado histórico ilustra bem o que aponta Budny:

Expressões figuradas como provérbios, expressões idiomáticas, combinações, colocações, e ditos, em geral são decodificadores naturais para o entendimento de crenças, costumes e convenções sociais; convenções que norteiam os usos das línguas e capacitam o falante a expressar e a entender pensamentos, emoções e visões de mundo como usuário daquele sistema de comunicação. (Budny, 2020, p. 308)

Por conta disso, segundo Xatara e Succi (2008), Els “ou outro fraseologismo qualquer” também são marcas evidentes da fluência e “do domínio dos falantes sobre as línguas, já que o seu uso requer competência lexical e cultural”. Ao utilizarem incorretamente ou não serem capazes de compreender as Els, os falantes de uma língua, e também os tradutores, correm o risco de se passarem por “falantes ingênuos” (Fillmore, 1979), ou mesmo “tradutores ingênuos” (Tagnin, 2002), que ainda não vivenciaram ou não foram suficientemente expostos ao idioma para conhecer seus sentidos socialmente convencionalizados, para além do significado de seus componentes.

Assim, apesar de sua ampla utilização e integração à linguagem, as Els são um notável desafio para a tradução e para o aprendizado de línguas (Tagnin, 2005; Simão, 2011; Termignoni e Finatto, 2017; Silva e Teixeira, 2021; Seco e Parreira, 2025), tanto pela opacidade e complexidade semântica de seu sentido quanto pelo fato de seu aprendizado e aquisição, em geral, ocorrerem por meio do contato reiterado com seu uso em contexto (Mattos, 2003; Termignoni e Finatto, 2017).

As Els também apresentam desafios para o PLN (Sag *et al.*, 2002), que impactam diretamente as aplicações linguísticas. Nos últimos dez anos, a IA, o PLN, os *Large Language*

Models - LLMs (grandes modelos de língua), as redes neurais artificiais, o aprendizado de máquina (*machine learning*) e o aprendizado profundo (*deep learning*) apresentaram grandes avanços e vêm revolucionando o mercado e as tecnologias da linguagem e da tradução, inclusive no que tange às Els. Em testes quantitativos recentes, envolvendo vários recursos linguísticos e línguas além do inglês, *chatbots* de IA como o ChatGPT (OpenAI) e Gemini (Google) foram capazes de identificar corretamente o uso de linguagem figurada em quase 100% das vezes (Alhaydar, 2025), sendo inclusive capazes de produzir traduções não literais para as Els. Entretanto, essa capacidade ainda parece estar relativamente restrita ao uso de explicitações, explicações e paráfrases (Raunak *et al.*, 2023). Outras vezes, os sistemas identificam a idiomaticidade, mas nem sempre conseguem recuperá-la no texto traduzido (Rebechi; Marcon; Faller, 2025).

Uma maneira de solucionar as dificuldades enfrentadas por tradutores, aprendizes, falantes de idiomas e pesquisadores que trabalham com as Els poderia ser a consulta a materiais de referência, como os tradicionais dicionários, glossários e enciclopédias. Embora esses materiais existam, eles são em grande maioria semasiológicos, ou seja, partem da expressão para o significado, pressupondo um conhecimento prévio da El que se busca. Esse tipo de material é irremediavelmente ineficiente para um tradutor ou aprendiz de língua estrangeira, que busca uma possível equivalência para transmitir algum pensamento ou sentido com o qual depara e deseja expressar em outra língua. Esse é o percurso de codificação do sentido, abordado pelas obras chamadas onomasiológicas, explicadas a seguir.

2.3 Materiais de consulta onomasiológicos

Dicionários, vocabulários e glossários **onomasiológicos** são materiais lexicográficos que partem do significado em direção ao significante, e não o contrário, como é comumente feito na grande maioria dos repertórios (percurso chamado de **semasiológico**). Em *Dicionários: uma pequena introdução à Lexicografia*, Welker (2004) estabelece que:

O dicionário alfabético é um dicionário semasiológico (do grego *semasía* – “significado”); ele vai da forma, do lexema, ao significado. No onomasiológico (do grego *onomasía* – “termo”), o movimento é o oposto: parte-se de conceitos para encontrar signos (cf. Baldinger 1960: 523). Embora o dicionário

alfabético seja o mais comum e muito antigo, o onomasiológico também tem uma longa tradição (Welker, 2004 *apud* Moreira de Oliveira, 2022, p. 47).

Um dicionário onomasiológico de Els permitiria, então, que o consulente partisse do sentido que quer expressar para chegar a alternativas de Els que se adequem à sua necessidade. Por exemplo, um tradutor de PB>EN pode encontrar a expressão “chutar o balde”, que significa “desistir de algo, geralmente com raiva e frustração, cansaço ou decepção”. Caso não conheça de antemão uma EI equivalente, a pessoa poderia recorrer à explicitação ou paráfrase, ou mesmo omitir a EI na língua de chegada, mas o texto perderia em idiomatidade. Entretanto, se pudesse realizar uma busca partindo do sentido para a expressão, poderia pesquisar por “desistência”, “impaciência”, “fúria” e encontrar expressões potencialmente equivalentes, como “*throw everything up in the air*”, “*go ballistic*”, “*hit the ceiling*” e então selecionar aquela que mais se adequasse à sua necessidade contextual. Se, por outro lado, fizesse uma tradução literal dos elementos da EI, buscando uma equivalência com base na tradução *ipsis litteris* dos elementos da EI, chegaria a “*kick the bucket*”, EI existente na língua inglesa, mas cujo significado é bem diverso: “morrer”.

Apesar de menos comuns do que dicionários semasiológicos, alguns dicionários e outros tipos de materiais de consulta onomasiológicos já foram produzidos no Brasil com temas diversos, envolvendo outras línguas, como, por exemplo, *Dicionário analógico da língua portuguesa: Tesouro de Vocábulos e Frases da Língua Portuguesa* (Spitzer, 1936), *Dicionário onomasiológico de expressões idiomáticas usuais na língua portuguesa do Brasil* (Riva, 2009), *Plataforma Kuhl pei: proposta de um modelo de dicionário terminológico onomasiológico multilíngue para crianças, Português – Arara, Kadiwéu, Karitiana, Parintintin, Xavante, Zoró* (Gava, 2012) e *Dicionário onomasiológico de expressões cromáticas da fauna e flora* (Martins, 2013).

Para a temática da alimentação, apesar de ser uma área que carrega marcas culturais muito fortes, ainda não há, salvo engano, dicionário onomasiológico multilíngue de Els publicado abrangendo esse tópico. Esse é o objetivo geral do projeto de pesquisa ao qual o trabalho aqui relatado se filia, explicado brevemente a seguir para contextualizar a importância de se desenvolver um sistema de classificadores semânticos como o que abordamos neste estudo.

3 Classificação semântica de expressões idiomáticas: padronizando anotações

Conforme mencionado anteriormente, o Projeto visa desenvolver uma plataforma multilíngue online, interativa, interrelacionada, de acesso livre e de inventário aberto, para o registro e disponibilização de Els contendo palavras do campo semântico da alimentação que permita catalogar, consultar e estudar essas Els de uma perspectiva onomasiológica. Inicialmente, estão sendo trabalhadas Els em português, inglês, espanhol, chinês e italiano, com possibilidade de expansão futura para outros idiomas.

A interface a ser desenvolvida permitirá várias opções de busca além da semasiológica, como, por exemplo, a busca por idioma, significado, campo semântico, grupo de alimento, categoria gramatical, estrutura sintática, grau de fixidez e idiomaticidade, variações dialetais, entre outros. Isso visa facilitar a compreensão e utilização de Els por tradutores, falantes e aprendizes de L1 e L2, além de permitir a utilização das informações geradas através do cruzamento de dados entre idiomas para diversas aplicações computacionais e educacionais futuras, por professores e pesquisadores dessas e de áreas correlatas, como PLN, Linguística Computacional e IA.

Para a realização deste Projeto, estipulou-se que um dos primeiros passos seria elaborar um sistema adequado e abrangente de anotação dos significados, temáticas e categorias semânticas relacionados a essas Els, além de coletar exemplos de uso reais. Esses descritores serviriam para identificar o sentido das Els sob várias perspectivas, facilitando sua localização pelos consulentes. Por exemplo, a EI “ALGO SER mamão com açúcar” está relacionada a conceitos como “facilidade”, “simplicidade” e “despreocupação”; “pagar o pato”, por sua vez, poderia ser associada a “culpa”, “responsabilidade” e “injustiça”. Isso possibilitaria, segundo a hipótese levantada pelo Projeto, realizar o cruzamento posterior dos dados de anotação das Els no âmbito de determinado idioma e no contraste entre idiomas distintos, permitindo o agrupamento, de forma (semi)automática, de Els equivalentes interlinguais e intralinguais, por meio da observação de descritores comuns.

Entendeu-se que a lista de descritores não poderia, por um lado, ser extensa demais, a ponto de resultar em categorias muito específicas e com poucas repetições entre as línguas. Isso poderia dificultar a identificação de coincidências semânticas entre as línguas, além de inviabilizar etapas posteriores de anotação humana, já que, quanto mais categorias disponíveis,

maiores as chances de erro e discordância entre anotadores. Por outro lado, tampouco poderia ser muito reduzida, para não acabar agrupando Els consideravelmente diferentes, como “perguntar se macaco quer banana”, que significa “fazer uma pergunta óbvia”, e “bananeira que já deu cacho”, que designa “alguém ou alguma coisa que já não tem futuro”, sob um mesmo descritor, por exemplo, “expectativa / probabilidade”. Em outras palavras, embora ambas as Els expressem, de alguma forma, os conceitos de “expectativa / probabilidade”, são diferentes o bastante para não poderem ser agrupadas sob a mesma etiqueta, especialmente quando se espera que essa etiqueta indique possíveis equivalentes em outras línguas.

Com isso em mente, foram feitos alguns experimentos de classificação de Els, inicialmente com foco nas que contêm palavras do campo semântico “frutas”, no par de línguas PB-EN, com o objetivo de: i) testar e refinar uma lista de descritores eficiente e replicável, que não fosse demasiadamente ampla ou concisa; ii) chegar a uma metodologia de classificação replicável, a ser aplicada às demais Els do banco de dados.

4 Metodologia

Com o objetivo de elaborar uma lista de classificadores semânticos aplicável à totalidade das Els do banco de dados em etapa futura, o primeiro passo deste estudo foi selecionar arbitrariamente um subconjunto das Els, em PB e EN, para um experimento inicial. Tal seleção, feita manualmente, foi uma estratégia metodológica para tentar contornar a dificuldade de classificar semanticamente a grande quantidade de Els até então coletadas nessas duas línguas – aproximadamente 700 em PB e 800 em EN –, permitindo que o estudo pudesse ser conduzido de maneira mais eficiente e organizada. O ponto de partida escolhido foram as Els contendo pelo menos uma palavra do campo semântico “frutas” em sua composição, a começar pelas Els assim classificadas no trabalho de Moreira de Oliveira (2022), em que foram estabelecidos campos semânticos relacionados aos alimentos (como *frutas*) e a seu preparo e consumo, como *utensílios* (p. ex. “toda panela tem sua tampa”), *métodos de cozimento* (p. ex. “a batata de ALGUÉM estar assando”), *bebidas* (p. ex. “tomar chá de cadeira”) etc.

Separadas as Els classificadas com a etiqueta *fruta* em PB e EN presentes no banco de dados atualizado, a etapa seguinte consistiu em elaborar uma lista concisa de **campos**

semânticos relacionados a seus possíveis significados, que foram selecionados a partir das seguintes fontes:

1. os classificadores únicos propostos por Moreira de Oliveira (2022), num total de 49 (PB);
2. a lista de “*Emotions*” utilizada na classificação refinada de emoções da base de dados do *GoEmotions*² (Demszky *et al.*, 2020), com 28 itens (EN);
3. uma lista de “emoções humanas” elaborada pelo ChatGPT, com 47 itens (PB, provavelmente traduzidas do EN);
4. os “*Themes*” (temas) utilizados pelo *Oxford Dictionary of Idioms* (Ayto, 2020), com 81 itens (EN).

Os campos semânticos apresentados pelas fontes acima foram analisados, comparados e consolidados para a elaboração de uma lista abrangente, mas passível de ser usada sistematicamente por várias pessoas com níveis de formação diversos e em várias línguas.

Inicialmente, todo o grupo de pesquisa se reuniu de maneira presencial e discutiu algumas das expressões, qual sentido elas exprimiam e que critérios usar para escolher, se possível, apenas um descritor da lista consolidada para caracterizá-las. O contato inicial com a tarefa foi feito para que os pesquisadores se familiarizassem com a metodologia e pudessem replicá-la depois, individualmente.

Em seguida, cinco pesquisadores do grupo³, classificaram a lista de Els em PB e EN contendo palavras do campo semântico “frutas”. As Els foram dispostas em uma planilha, separadas por idioma e agrupadas por frutas – p. ex., todas as Els contendo “banana” foram agrupadas. Na coluna ao lado de cada EI estavam definições e explicações retiradas de referências bibliográficas consultadas e informações encontradas na internet, como mostra a vista parcial da planilha na Figura 1. A tarefa era escolher um dos 90 classificadores da lista e, nos casos em que apenas um não fosse suficiente, ou em que a lista não contivesse um classificador considerado o mais adequado para a EI em questão, os anotadores usariam a coluna mais à direita para propor outro(s), além de anotar possíveis observações.

² <https://github.com/google-research/google-research/tree/master/goemotions>

³ três doutoras, uma bacharela em Letras - Línguas Estrangeiras Aplicadas e graduanda em Tradução Inglês, e um graduando em Letras - Língua e Literatura Inglesa.

Quadro 1 – Trecho do instrumento inicial de classificação semântica de Els respondido por uma anotadora.

Expressão idiomática	Definição / explicação	Classificador único (se fosse 1 só)	Outros possíveis / necessários, observações
ALGO SER mamão com açúcar	Coisa muito fácil. (https://tinyurl.com/c67a6p9e)	facilidade	
ALGO SER a cereja do bolo	Expressão usada para falar sobre aquilo que serve como finalização perfeita para algo ou algum evento. A cereja do bolo da festa foi a linda dança dos recém-casados. (https://tinyurl.com/2mptnt2u)	perfeição	conclusão
DESCASCAR um/o abacaxi	Resolver um pepino, descascar um abacaxi - resolver um problema, dar um jeito em uma situação problemática. (https://tinyurl.com/knhrvzmmh)	perturbação	problema
ALGUÉM SER um doc(e)/inho (de coco)	Expressão utilizada para dizer que alguém é um amorzinho, uma gracinha. Você é um docinho de coco. (https://tinyurl.com/72xmv7d6)	aparência	
ALGUÉM ENFIAR o pé na jaca	A expressão enfiar o pé na jaca está dicionarizada com o significado de beber muito, além do limite, ou seja, tomar um porre. Por extensão, é usada também para indicar qualquer tipo de excesso ou exagero, quer a nível alimentar, quer a nível comportamental. (https://tinyurl.com/5yja2wmb)	exagero	

Fonte: autoria própria.

Os resultados individuais foram então comparados e discutidos em reunião, momento em que cada pesquisador(a) justificava suas escolhas. A partir do resultado desse debate, a lista foi mais uma vez refinada. Por fim, foi feito o cruzamento dos dados para verificar o quanto a metodologia se mostrou eficaz na identificação de possíveis equivalentes interlinguais e de relações de sinonímia intralingual entre as Els classificadas.

5 Apresentação e discussão dos resultados

5.1 A lista de classificadores

Depois de comparar as quatro listas utilizadas como ponto de partida para a elaboração de uma lista de classificadores única para aplicar às Els do banco de dados, a começar pelas que contêm pelo menos o nome de uma fruta, observamos algumas semelhanças e várias diferenças entre elas. A lista proveniente do trabalho de Moreira de Oliveira (2022), com 49 itens em português brasileiro (lista 1), assim como a extraída do dicionário Oxford de expressões idiomáticas (lista 4), misturam vários tipos de descritores, como mostram os exemplos do Quadro 2.

Quadro 2 - Exemplos de descritores presentes em Moreira de Oliveira (2022) e Ayto (2020).

tipo de classificador	exemplos da lista 1	exemplos da lista 4
A. emoções	desprezo, desconfiança, insegurança	<i>happiness, embarrassment, disappointment</i>
B. atitudes / traços de personalidade	arrogância, avareza	<i>courage, hypocrisy, self-interest</i>
C. situações / acontecimentos	fofoca, oportunismo, impunidade	<i>death, pregnancy, crisis</i>
D. campos semânticos temáticos	partes do corpo, aparência, trabalho	<i>weather, family, clothes</i>

Fonte: autoria própria.

Por outro lado, a lista proveniente do *GoEmotions* (lista 2) e a elaborada pelo ChatGPT (lista 3), com 28 (EN) e 47 (PB) itens, respectivamente, contêm apenas emoções. Apesar de a lista 2 ser bastante curta, se comparada às outras três, elenca emoções que não identificamos em nenhuma das outras listas, como “*amusement*”, “*caring*”, “*nervousness*”, “*grief*”, “*optimism*” e “*realization*”, além da etiqueta “*neutral*”, talvez usada para expressar a ausência de emoções. Vale notar que a lista traz o classificador “*pride*”, que coincide apenas parcialmente com “orgulho”, da lista 3, já que o primeiro costuma ter uma prosódia semântica mais positiva, ao passo que “orgulho” poderia ser representado também por “*arrogance*” (“arrogância”), da lista 1, e “*boastfulness & conceit*”, da lista 4.

Também saltou aos olhos, na comparação das quatro listas, o uso de classificadores bastante semelhantes, que poderiam ser considerados sinônimos, tais como i) “alegria”, “felicidade” e “contentamento”; ii) “raiva” e “ódio”; iii) “culpa”, “remorso” e “arrependimento”; iv) “desgosto” e “frustração”; e v) “empolgação” e “excitação”, na lista 3. Ou vi) “*doubt & uncertainty*” e “*indecision & prevarication*”; e vii) “expense”, “debt” e “money, wealth & prosperity”, na lista 4. Após reuniões e debates, chegamos a uma lista de 90 classificadores, utilizada no experimento de classificação descrito no item a seguir.

5.2 O experimento de classificação

As Els classificadas manualmente como contendo uma palavra do campo semântico “fruta” no banco de dados totalizaram 60 em PB e 87 em EN. Foram filtradas na planilha usando esse critério e separadas para a realização do experimento de classificação semântica. Após serem dispostas no documento exibido no Quadro 1 acima, foram enviadas aos participantes do projeto, juntamente com a lista de 90 classificadores em português e inglês (foi usada tradução automática minimamente editada da lista de português para criar a lista do inglês). Cinco pesquisadores finalizaram a tarefa. Os dados foram reunidos numa planilha para comparação (vide Figura 1).

Figura 1 - Vista parcial da planilha de coleta e comparação dos dados de anotação PB.

	A	B	C	D	E	F	G
1	Expressão idiomática	Definição / explicação	Anotador/a 1	Anotador/a 2	Anotador/a 3	Anotador/a 4	Anotador/a 5
113	ALGUÉM/ ALGO SER laranja de amostra	Uma coisa muito bonita de se ver. https://tinyurl.com/3swpjhtm	Aparência	aparência	aparência (positivo)	Elogio	Aparência?
114	Mais perdido que cebola em salada de fruta	desorientado, deslocado	Improbabilidade	confusão	confusão (mental)?	Confusão	Trapalhada
115	Me caiu / de cair os butiá do bolso	É uma expressão regionalista típica do Rio Grande do Sul. Usa-se quando a pessoa quer dizer que está impressionada, assustada ou estupefata.	Surpresa	surpresa	surpresa	Surpresa	Surpresa
116	PERGUNTAR se macaco quer banana	Fazer pergunta óbvia, afirmar o óbvio	Obviedade	obviedade (previsibilidade)	obviedade	Obviedade	Obviedade
117	Na sombra da bananeira	em situação muito favorável	Felicidade	plenitude	facilidade	Plenitude	Facilidade
118	ALGUÉM SER o cão chupando manga	Situação improvável, estranha, difícil ou fora do comum; algo ou alguém desprezível, vil ou indigno de admiração. Em algumas regiões do país, é usada para descrever alguém extremamente feio, horrendo ou desagradável à vista.	Aparência	aparência	aparência (neg)	Insulto	Aparência

Fonte: autoria própria.

Seis expressões em PB não continham o nome de uma *fruta*, mas de um *fruto* (isto é, “parte do vegetal oriunda do desenvolvimento do ovário e que tem como função proteger e cobrir as sementes”), e foram descartadas desse recorte: 2 Els com “azeitona” e 4 com “quiabo”. Isso reduziu a lista de Els do português para 54.

Observamos que apenas 60 (dois terços) do total dos 90 descritores disponíveis foram utilizados. Avaliamos que isso pode ser uma limitação imposta pelo recorte feito, ao termos decidido trabalhar apenas com Els contendo ao menos uma lexia da categoria *fruta*. Desses 60, cerca de metade (32) foram usados apenas uma vez. Os mais frequentes em PB foram *aparência* (21 ocorrências), como em “maracujá de gaveta”; *facilidade* (14 oc.), como em “mamão com açúcar”; e *exagero* (11 oc), como em “muita banana para pouco tostão”. No inglês, essa contabilização ficou prejudicada pelo fato de que alguns anotadores usaram

classificadores em português e outros em inglês, impossibilitando a contagem de forma automatizada – uma questão a ser reforçada nos próximos experimentos.

5.3 Contrastando os classificadores

Ao cruzar os dados provenientes da anotação, consideramos apenas os descritores que foram usados por pelo menos três dos cinco anotadores para uma mesma EI (vide, por exemplo, as EIs da Figura 1, com exceção da linha 117). Observamos, então, sua eficiência para a identificação de possíveis equivalentes interlinguais e também para reunir EIs de sentido aproximado dentro de cada uma das línguas (relações intralinguais). Os exemplos a seguir ilustram esse procedimento.

A EI do PB “PERGUNTAR se macaco quer banana” foi classificada pelos 5 anotadores com *obviedade*; em EN, “(as) sure as God made little (green) apples” teve o mesmo classificador associado por três dos cinco anotadores - um bom exemplo de candidato a equivalente. O mesmo se deu com as EIs em EN “cherry pie” e “low-hanging fruit”, ambas classificadas como “facilidade” por quatro avaliadores, e por isso consideradas como possíveis equivalentes para a EI em PB “ALGO SER mamão com açúcar”.

De uma perspectiva intralingual, as EIs nas linhas 113 e 118 da Figura 1, ambas classificadas por mais de três anotadores como *aparência*, não podem ser consideradas exatamente sinônimas em PB, já que “ALGUÉM / ALGO SER laranja de amostra” tem uma prosódia semântica positiva, ao passo que “ALGUÉM SER o cão chupando manga” tem prosódia negativa. Mas ambas se relacionam, de fato, com a *aparência* - isto é, fazem parte desse campo semântico mais amplo e são, portanto, relacionadas. Outras EIs do PB classificadas com essa etiqueta foram “ALGUÉM TER uma pele de pêssego”, “ALGUÉM SER / PARECER (um) maracujá de gaveta”.

Alguns dos classificadores que resultaram em EIs equivalentes PB-EN, incluindo os exemplos mencionados, foram:

- I) *facilidade*: ALGO SER mamão com açúcar -> *low-hanging fruit*; *cherry pie*;
- II) *exagero*: ALGUÉM ENFIAR o pé na jaca, dar uma jacada (variante) -> *belt the grape*;
- III) *superação*: se/quando a vida DAR A ALGUÉM um limão/limões, FAZER (uma) limonada, fazer de um limão uma limonada (variante) -> *if life gives you lemons, make lemonade*;

IV) *estima*: ALGUÉM SER um doc(e/inho) (de coco) -> *peach*; *peacherino*; *peachy*; *peachy-keen*;

V) *constrangimento*: FICAR vermelha/o como/que nem (um) tomate -> *as red as cherry*;

VI) *produtividade*: ALGO dar/render (bons) frutos (sinônimo) -> *bear fruit*;

Alguns descritores foram usados para classificar somente uma EI cada, resultando na impossibilidade de cruzamento de dados entre as EIs do banco de dados já classificadas, o que não impede que levem a possíveis equivalentes e EIs relacionadas no futuro:

I) *egoísmo*: “em tempos de murici, cada um por si”;

II) *reclamação*: “ALGUÉM CHORAR (as/POSSESSIVO) pitangas”.

Devido ao fato de terem sido classificados de maneiras diversas e não apresentarem maioria, os classificadores falharam em apresentar as seguintes EIs como sinônimas, variantes ou equivalentes:

I) *perturbação, contaminação, problema*: “pomo da discórdia” -> “*apple of discord*”; “*bad apple*”; “*rotten apple*”;

II) *dificuldade, confusão, complexidade, perturbação, trapalhada*: “ALGO SER uma bananosa” -> “*BE a lemon*”;

III) *perfeição, estima*: “ALGUÉM SER uma uva” -> “*BE a plum*”.

As EIs com significados equivalentes “ALGO SER uma bananosa” e “*SOMETHING BE a lemon*” relacionam-se semanticamente a “ALGO ser um abacaxi” (e a “ALGO SER um pepino”, que não fez parte deste recorte do banco de dados).

5.4 Discussão

O fato de os anotadores terem usado classificadores em português e em inglês para classificar as EI em EN dificultou enormemente a tabulação dos dados, bem como o fato de terem feito observações e sugestões (como o uso da interrogação) nas colunas de classificação, em vez de reservarem essas sugestões / alterações para a coluna apropriada (vide Figura 2). Isso nos levou a crer que um formulário com um menu suspenso para a escolha do descritor talvez seja uma forma mais eficiente de coletar as classificações de anotadores do que permitir ou solicitar que digitem o classificador.

Figura 2 - Vista parcial da planilha de coleta e comparação dos dados de anotação EN.

	A	B	C	D	E	F	G
9	be in apple-pie order	Neat and tidy, and everything is where it should be.	Excellence perfection	perfeição	perfection	Perfection?	Perfeição
10	apple-sauce	bunkum, nonsense	foolishness	Trapalhada?	confusion (mental)	Deception	insensatez
11	(one) only gets one bite at the apple	(LAW) A single opportunity to litigate a civil case against another party. (Once a case has been resolved by a judge, a plaintiff or prosecutor cannot pursue the same case against the same party again in the future.)	opportunity	dificuldade (Excepcionalidad e??)	opportunity	Opportunity	Oportunidade
12	bad apple	bad person	(evilness)	contaminação	problem (evil)	Aversion ?	aversão

Fonte: autoria própria.

Por outro lado, um menu *drop down* com 90 itens nos pareceu um tanto extenso, o que nos levou a fazer uma nova varredura na lista de classificadores, com o objetivo de identificar descritores que pudessem ser agrupados. Encontramos os seguintes (em negrito, os que foram mantidos):

- *aparência, beleza, velhice*
- *diferença, distinção, **excepcionalidade***
- *constrangimento, vergonha*
- *contrariedade, **dificuldade**, perturbação, problema*
- *êxito, solução, **superação***
- *excelência, **perfeição**, plenitude*
- *homossexualidade, **sexualidade**, virgindade*
- *aversão, desdém, **desprezo**, indiferença, rejeição*
- *esgotamento, **exaustão***
- *desonestidade, mau-caratismo*
- *agrado, **estima**, favoritismo, predileção*

Além disso, dois classificadores da lista inicial foram substituídos por palavras que parecem expressar mais claramente o sentido pretendido: *contaminação* foi substituída por *influência* (usada em Els como “SOMEONE BE a rotten apple” e “ALGUÉM/ALGO SER o pomo da discórdia”) e *engodo* substituída por *enganação* (como em “dead sea fruit” e “ALGUÉM ser [ARTIGO] laranja”). Disso resultou uma nova lista, com 63 itens:

1. aparência	22. exaustão	43. maciez
2. arrogância	23. excepcionalidade	44. mau humor
3. atração	24. exibicionismo	45. naturalidade
4. bajulação	25. expediente	46. negativismo
5. barganha	26. facilidade	47. obviedade
6. coibição	27. felicidade	48. oportunidade
7. confusão	28. frustração	49. parte do corpo
8. constrangimento	29. impaciência	50. perfeição
9. degradação	30. impossibilidade	51. perigo
10. desconfiança	31. improbabilidade	52. previsibilidade
11. desdobramento	32. improdutividade	53. produtividade
12. desonestidade	33. incompetência	54. reclamação
13. desprezo	34. indignação	55. ressentimento
14. dificuldade	35. influência	56. seletividade
15. egoísmo	36. ingenuidade	57. sexualidade
16. embriaguez	37. insanidade	58. simploriedade
17. enganação	38. insensatez	59. solução
18. esperteza	39. insignificância	60. supervisão
19. estima	40. inutilidade	61. surpresa
20. evasão	41. lerdeza	62. trapalhada
21. exagero	42. liderança	63. tristeza

Outro aspecto observado a partir dos resultados do experimento é o fato de que, ao oferecermos uma definição ou explicação retirada de outras fontes, perdemos a oportunidade de saber se a pessoa que estava anotando conhecia ou não a EI. Isto é, se sua classificação se baseava em conhecimento próprio, advindo do contato com a EI no uso da língua, ou na informação de certa forma “imposta” por essas definições / explicações apresentadas, que seria o único critério utilizado para escolher o classificador semântico, já que esse conhecimento não provinha de experiências reais de uso da língua. Isso nos levou a pensar que, em experimentos futuros, talvez fosse mais proveitoso oferecer contextos autênticos de uso, provenientes de corpora, em vez de definições e explicações prontas, para simular esse contato inicial com uma EI desconhecida e a possibilidade (ou não) de a/o anotador/a depreender seu sentido a partir desses usos.

6 Considerações finais

Os resultados obtidos neste estudo evidenciam tanto o potencial quanto as limitações da utilização de classificadores semânticos atribuídos por anotadores humanos para a identificação de equivalentes interlinguais e intralinguais para Els. O baixo índice de consenso observado confirma a complexidade inerente à caracterização semântica desse tipo de unidade lexical e aponta para a necessidade de uma metodologia mais consistente e de instrumentos de anotação mais claros e uniformes.

Apesar dessas dificuldades, a experiência relatada permitiu diversos avanços. A sistematização de uma lista inicial de classificadores, ainda que sujeita a alterações, mostrou-se viável como ponto de partida para futuros experimentos. Além disso, os exemplos de coincidência entre classificadores em PB e EN sugerem que há espaço para a consolidação de descritores que favoreçam a identificação (semi)automática de equivalentes idiomáticos.

Em termos práticos, os resultados sinalizam a necessidade de equilibrar sistematicidade e flexibilidade: por um lado, é imprescindível dispor de descritores suficientemente claros e replicáveis; por outro, não se pode ignorar a multiplicidade de leituras possíveis de uma mesma expressão idiomática. Assim, este trabalho representa um passo inicial no caminho para “descascar esse abacaxi”, ou seja, construir uma metodologia que permita mapear, organizar e disponibilizar de forma eficaz as expressões idiomáticas em diferentes idiomas para consulta onomasiológica.

Este estudo destaca que, caso esse percurso metodológico de anotação humana das Els continue a ser adotado, é necessário fornecer instruções mais específicas para quem vai fazer o trabalho, a fim de mitigar variáveis que influenciam a forma como o teste é respondido e tabulado. Talvez seja produtivo realizar algum treinamento prévio com os participantes, apresentando exemplos de classificação e como resolver problemas e indecisões. Além disso, julgamos importante coletar dados adicionais sobre a familiaridade dos participantes com as Els, como, por exemplo, se já as conheciam, ou pedir uma autoavaliação do quanto acreditam ter compreendido seu significado a partir dos contextos reais de uso oferecidos, uma vez que são fatores que podem influenciar a escolha de descritores e, portanto, a concordância geral do grupo na anotação dos classificadores.

Este estudo também nos leva a questionar o método em si. É realmente necessário e eficiente escolher um descritor semântico principal para cada EI, como nossa hipótese inicial sugeria, para que seja possível cruzar os dados? Considerando a multiplicidade de interpretações e visões de cada indivíduo, como aconteceu com “ALGO dar/render pano pra manga”, classificado como *produtividade*, *desdobramento (2x)*, *complexidade* e *reclamação*, não seria melhor permitir o uso de uma quantidade maior de descritores para cada EI? Se é difícil haver consenso entre estudantes e professores de Letras quanto a um descritor semântico de uma EI, é de se supor que futuros usuários da plataforma também tenham dificuldade para selecionar um descritor que identifique adequadamente o campo semântico do significado que busca.

A partir dessas experiências, estão sendo feitos ajustes à metodologia e à sua aplicação, como também a reavaliação das reais necessidades e possibilidades que envolvam a criação de um sistema informatizado de anotação de classificadores. Espera-se que, por meio dessas reflexões, sejam promovidos novos direcionamentos para a criação de materiais lexicográficos onomasiológicos voltados às EIs, e novos caminhos sejam indicados para o aprimoramento de tecnologias linguísticas em ambientes digitais que tratem desse tópico fundamental.

Referências

ADEWUMI, T; VADOODI, R; TRIPATHY, A; NIKOLAIDOU, K.; LIWICKI, F.; LIWICKI, M. Potential Idiomatic Expression (PIE)-English: Corpus for Classes of Idioms. *In: Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, Marseille, 20-25, 2022, p. 689–696. <https://doi.org/10.48550/arXiv.2105.03280>.

ALHAYDAR, A. NLP and Figurative Language: A Quantitative Study Exploring the Competence of AI-Powered Bots in Understanding Metaphors and Idioms. *Journal of AI*, v. 9, n. 1, p. 32-55, 2025. <https://doi.org/10.61969/jai.1613027>.

AYTO, J. *The Oxford Dictionary of Idioms*. Oxford: Oxford University Press, 2020. <https://doi.org/10.1093/acref/9780198845621.001.0001>.

BAKER, M. *In Other Words: A Coursebook on Translation*. London: Routledge, 1992.

BHALA, J. *I’m not hanging noodles on your ears*. Washington, D.C.: National Geographic Society, 2009.

BUDNY, R. Unidades fraseológicas com zoônimos: um olhar pela janela da influência cultural. **Linguagem em (Dis)curso**, Santa Catarina, v. 20, n. 2, p. 307-319, 2020. <http://dx.doi.org/10.1590/1982-4017-200205-5419>.

CORPAS PASTOR, G. **Manual de fraseología española**. Madrid: Gredos, 1996.

COWIE, A. P. **Phraseology: theory, analysis, and applications**. Oxford: Oxford University Press, 1998.

COWIE, A. P.; MACKIN, R. **Oxford Dictionary of Current Idiomatic English**, Vol. I: Verbs with Prepositions and Particles. London: Oxford University Press, 1975.

DEMSZKY, D.; MOVSHOVITZ-ATTIAS, D.; KO, J.; COWEN, A.; NEMADE, G.; RAVI, S. GoEmotions: A Dataset of Fine-Grained Emotions. **ACL 2020**. <https://doi.org/10.48550/arXiv.2005.00547>.

FILLMORE, C. J. Innocence: A Second Idealization for Linguistics. **Annual Meeting of the Berkeley Linguistics Society**, [s. l.], 1979, p. 63–76.

GAVA, Á. A. **Plataforma Kuhi pei: proposta de um modelo de dicionário terminológico onomasiológico multilíngue para crianças, Português – Arara, Kadiwéu, Karitiana, Parintintin, Xavante, Zoró**. 2012. 329 f. Tese (Doutorado em Estudos Linguísticos). Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2012.

GRAMLEY, V.; GRAMLEY, S.; PÄTZOLD, K.-M. **A survey of modern English**. London: Routledge, 2003.

GRIES, Stefan Th. Phraseology and linguistic theory: A brief survey. In: GRANGER, S.; MEUNIER, F. (org.). **Phraseology: an interdisciplinary perspective**. Amsterdam: John Benjamins, 2008. p. 3-25.

GROSS, M. **Les limites de la phrase figée**. Langages, 90. Paris: Larousse, 1988.

JACOBSON, J.D. **Eatioms**. New York: Laurel, 1993.

KORKONTZELOS, I.; ZESCH, T.; ZANZOTTO, F. M.; BIEMANN, C. Semeval-2013 task 5: Evaluating phrasal semantics. In: Second Joint Conference on Lexical and Computational Semantics, Vol. 2: **Proceedings** of the 7th International Workshop on Semantic Evaluation, 2013, p. 39-47.

LACERDA, R. C., LACERDA, H. R. C., ABREU, E. S. **Dicionário de provérbios: francês, português, inglês**. São Paulo: Editora Unesp, 1999.

LEFFA, V. J. Aspectos externos e internos da aquisição lexical. In: LEFFA, V. J. (Org.). **As palavras e sua companhia: o léxico na aprendizagem**. Pelotas, 2000, v. 1, p. 15-44.

MARTINS, Sabrina de Cássia. **Dicionário onomasiológico de expressões cromáticas da fauna e flora**. 2013. 220 f. Dissertação (Mestrado). Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2013.

MATTOS, M. Fraseologia: conceitos e características para a identificação das locuções verbais. *Língua e Literatura*, v. 27, p. 271-300, 2003. <https://doi.org/10.11606/issn.2594-5963.lilit.2003.105469>.

MEL'ČUK, Igor. **Semantic description of lexical units in an explanatory combinatorial dictionary**. Amsterdam: John Benjamins, 1988.

MOREIRA DE OLIVEIRA, I. **Expressões idiomáticas com a temática alimentação: uma proposta de glossário português-inglês**. 2022. 137 f. Dissertação (Mestrado) — Universidade de Brasília, Instituto de Letras, Brasília, 2022.

MOREIRA DE OLIVEIRA, I.; TEIXEIRA, E. D. Expressões idiomáticas com a temática alimentação: um estudo contrastivo português-inglês. In: Razky, A.; Mejri, S.; Salvador, C. F. N. (org.). **A fraseologia: o paradoxo do universal e do idiomático / La phraséologie: le paradoxe de l'universel et de l'idiomatique**. Campinas: Pontes Editores, 2025, p. 195-236.

NEWMARK, P. **A Textbook of Translation**. Reino Unido: Prentice Hall, 1988.

ORTÍZ ALVAREZ, M. L. **Expressões idiomáticas do português do Brasil e do espanhol de Cuba: estudo contrastivo e implicações para o ensino de português como língua estrangeira**. 2000. 334 f. Tese (Doutorado em Linguística Aplicada) — IEL, Universidade Estadual de Campinas, Campinas, 2000.

OLIVEIRA, H. De onde surgiu a expressão “rei da cocada preta”? **Super Interessante**, 2017. Disponível em: <https://super.abril.com.br/coluna/oraculo/de-onde-surgiu-a-expressao-rei-da-cocada-preta/>. Acesso em: 11 jul. 2025.

PALMATIER, R. **Food: A Dictionary of Literal and Nonliteral Terms**. London: Greenwood Press, 2000.

PINNAVAIA, L. **Food and Drink Idioms in English: “A little bit more sugar and lots of spice”**. Cambridge: Cambridge Scholars Publishing, 2018.

PLANTIN, R. S. Gastronomismos linguísticos: um olhar sobre fraseologia e cultura. In: ORTÍZ ALVAREZ, M. L.; UNTERNBÄUMEN, E. H. (Org.) **Uma (re)visão da teoria e da pesquisa fraseológicas**. Campinas: Pontes, 2011, p. 249-275.

POLGUÈRE, Alain. From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, Oxford, v. 27, n. 4, p. 396-418, 2014.

RAMOS-NOGUEIRA, L. C. **La traducción de la fraseología en la obra de Carlos Ruiz Zafón en el par lingüístico español-portugués**. [s.l.] Universidad de Granada, 2017. Disponível em: <https://digibug.ugr.es/handle/10481/47400>.

RAUNAK, V.; MENEZES, A.; POST, M.; AWADALLA, H. H. Do GPTs Produce Less Literal Translations? *In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. Volume 2: Short Papers. Association for Computational Linguistics, 2023. p. 1041-1050.

REBECHI, R. R.; MARCON, N. O.; FALLER, G. A. Tradução automática e tradução humana de expressões multipalavras: descascando esse abacaxi. **ReVEL**, v. 23, n. 44, p. 346- 380, 2025.

REBECHI, R. R.; TRINDADE, E. Traduzir Metáfora não é Mamão com Açúcar: a Busca por Equivalentes de Botanomorfismos. **Polissema – Revista de Letras do ISCAP**, n. 21, p. 110-132, 2021. <https://doi.org/10.34630/polissema.vi21.4250>.

RIVA, H. C. **Dicionário onomasiológico de expressões idiomáticas usuais na língua portuguesa no Brasil**. 2009. 311 f. Tese (Doutorado em Estudos Linguísticos). Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2009.

SAG, I. A., BALDWIN, T., BOND, F., COPESTAKE, A., FLICKINGER, D. Multiword Expressions: A Pain in the Neck for NLP. *In: Third International Conference of Computational Linguistics and Intelligent Text Processing*, 17-23 fev. 2002, Cidade do México. **Proceedings** [...]. Alemanha: Springer-Verlag, 2002, p. 1-15. https://doi.org/10.1007/3-540-45715-1_1.

SECO, M. **Gastronomismos nas Expressões Idiomáticas do português do Brasil e seus correspondentes em francês da França**. 191 f. Dissertação (Mestrado em Estudos Linguísticos). Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2017.

SECO, M. **Proposta de uma base léxico-cultural bilíngue (FF-PB) online: expressões idiomáticas relacionadas a gastronomismos**. 2024. 340 f. Tese (Doutorado em Estudos Linguísticos). Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2024.

SECO, M.; PARREIRA, M. C. Gastronomismos: perspectivas fraseológica, tradutológica e pedagógica. In: Razky, A.; Mejri, S.; Salvador, C. F. N. (Orgs.). **A fraseologia: o paradoxo do universal e do idiomático / La phraséologie: le paradoxe de l'universel et de l'idiomatique**. Campinas: Pontes Editores, 2025. p. 169-193.

SILVA, J. M.; TEIXEIRA, E. D. Expressões idiomáticas com a temática alimentação: tradução e glossário de "Pepinos e Abobrinhas", de Márcio Alemão. **Tradterm**, v. 37, n. 2, p. 397-429, 2021. <https://doi.org/10.11606/issn.2317-9511.v37p397-429>.

SIMÃO, A. K. Estereótipos linguísticos: questões tradutórias derivadas do tratamento de fraseologismos em dicionários bilíngues. **Tradução & Comunicação: Revista Brasileira de**

Tradutores, v. 23, p. 21-78, 2011. Disponível em: <http://hdl.handle.net/11449/122283>. Acesso em: 11 out. 2020.

SPITZER, C. **Dicionário analógico da Língua Portuguesa: Tesouro de vocábulos e frases da Língua Portuguesa**. Rio de Janeiro: Editora Globo, 1936.

TAGNIN, S. E. O. A tradução dos idiomatismos culturais. **Trabalhos de lingüística aplicada**, nº 11, p. 43-52, 1988.

TAGNIN, S. E. O. Os corpora: instrumentos de auto-ajuda para o tradutor. **Cadernos de Tradução**, v. 1, n. 9, p. 191-219, 2002.

TAGNIN, S. E. O. **O jeito que a gente diz: combinações consagradas em inglês e português**. Barueri: Disal, 2005.

TERMIGNONI, S.; FINATTO, M. J. B. Sobre a importância de ensinar expressões idiomáticas. **Revista de Italianística**, n. 35, p. 112-124, 2017. Disponível em: <http://www.revistas.usp.br/italianistica/article/view/141895>. Acesso em: 06 jun. 2025.

WELKER, H. A. **Dicionários: uma pequena introdução à lexicografia**. 2a ed. rev. e ampl. Brasília: Thesaurus, 2004.

XATARA, C. M. Tipologia das expressões idiomáticas. **Alfa**, São Paulo, v. 42, p. 169-176, 1998.

XATARA, Claudia Maria; RIVA, Huéinton; RIOS, Tatiana. Tradução de idiomatismos. **Cadernos de Tradução**, v. 2, n. 8, p. 183-194, 2001.

XATARA, C. M.; SUCCI, T. M. Revisitando o conceito de provérbio. **Veredas On Line – Atemática**, Juiz de Fora, v. 1, p. 33-48, 2008.

As datas abaixo serão preenchidas pelo editor.

Recebido em: 29/09/2025

Aprovado em: 13.03.2026